Misc. Publ. V.C.R.C. (15) 1990.

APPLICATION OF MATHEMATICAL AND STATISTICAL TECHNIQUES TO STUDIES ON EPIDEMIOLOGY AND CONTROL OF LYMPHATIC FILARIASIS.

S. Subramanian., P. Vanamail., A.Manoharan & A. Srividya

Edited By

P.K. Rajagopalan

Vector Control Research Centre

Pondicherry 605 006

India.

0163

COMMUNITY HEALTH CELL

Application of Mathematical and Statistical Techniques to studies on Epidemiology and control of Lymphatic filariasis.

CONTENTS

1. Introduction. 1
1.1. Filariasis (extent of the problem). 1
2. Collection of parasitological data. 2
2.1. The need for a sampling design. 2
2.1.1. Principal steps involved in a sample survey. 2
2.1.2. Use of supplementary information. 4
2.1.3. Probability sampling methods. 5
2.1.4. The sample design for filariasis survey. 6
2.1.5. The Estimation of Sample Size. 7
2.1.6. Parameter Estimation. 8
2.2. Sampling design for clinical survey. 8
2.3. Frequency distribution of microfilaria counts. 10
2.3.1. Introduction. 10
2.3.2. Frequency Distribution. 11
2.3.3. Negative Binomial Distribution. 11
2.3.4. Method of Fitting Negative Binomial Distribution. 12
2.3.5. Frequency Distribution of Parasite in Human host. 13
2.3.6. Estimation of Excess Zeros in the Sample. 14
2.4. Estimation of the age specific rates of acquisition and loss of infection.. 14
2.4.1. Introduction. 14
2.4.2. Catalytic Models. 14
2.4.3. Application of Catalytic Models In Lymphatic Filariasis. 16
2.4.4. Calculation of the life-span of adult worm. 17
2.4.5. Calculation of gain of infection. 17
2.4.6. Calculation of corrected gain. 18
2.5. Estimation of risk of developing disease in relation to microfilaraemia status, and, true infection prevalence in a population. 19
2.5.1. Estimation of the proportion ever infected using Catalytic model. 20
2.5.2. Estimation of the risk of developing clinical Symptoms. 20
2.5.3. Estimation of true infection prevalence using the Probability of acquiring a female worm. 21
2.6. Spatial distribution of parasite carriers. 23
2.6.1. Analysis . 24
3. Entomological data. 25
3.1. Introduction. 25
3.2. Sampling the larval population. 25
3.2.1. The Sampling design. 25
3.3. Sampling the adult population. 26
3.3.1. Estimation of entomological Parameters. 27
3.3.1.1. The survival rate of vector. 27
3.3.1.2. Distribution of parasite in the vector mosquitoes. 28
4. Concluding Remarks. 29
5. References. 30

01656

OH100

1. INTRODUCTION

The basic reasoning behind the epidemiological studies of any disease is to give guidelines to planners about the importance of the disease in an area and to monitor and evaluate changes in the endemicity due to control programmes. Since most of the disease control operations are long term programme and implemented in a phased manner, information in every phase of the programme (whether it is a planning, preparatory phase, evaluation or implementation phase) need to be processed, quantified and rational feedbacks need to be given for the improvement if any required. The disease could be quantified by estimating different parameters. These parameters can be estimated either by direct crude methods or by use of sophisticated mathematical formulations. Estimates obtained by either by crude method or by mathematical model can be used to evaluate the effectiveness of control programmes during different phases. Therefore, adequate knowledge on statistical/mathematical principles of designing studies, collection of correct and adequate sample and a rational analysis, etc. are highly essential not only for carrying out the study but also for drawing meaningful conclusions.

Though the theory of mathematics developed for biological applications are extensive and often so abstruse, its implications are no longer accessible to many biologists. Therefore, it is an essential task for applied mathematicians to maintain a communication link between the mathematical theory and biological reality. Such a communication link, if rightly understood, is expected to result in far reaching practical consequences. In this training module the authors tried to develop a link between mathematical theory and biological reality in relation to an important vector borne disease filariasis.

1.1. FILARIASIS (EXTENT OF THE PROBLEM):

Filariasis is a major health problem in tropical countries of the world. Though mortality due to filariasis is negligible, there is a high degree of morbidity due to its acute and chronic manifestations. According to latest WHO estimates 2677 million people are living in countries where the disease is endemic, 905 million people are at the risk of infection (living in areas where transmission is active) and 90.2 million people are infected[1-4]. Over two thirds of lymphatic filariasis cases of the world are contributed by India, China and Indonesia. Lymphatic filariasis is an age old disease recorded in India as early as 6th century B.C. Though both *bancroftian* and *brugian* forms of lymphatic filariasis are prevalent in India, bancroftian is found to be the commonest, accounting for 98% of the filarial infection. A continuously increasing trend was observed from 1962 (5.3 million mf carriers) to 1989 (27 million mf carriers). Published results have shown that 377 million people in India are exposed to the risk of *W.bancrofti* and *B.malayi* infection with an estimated 27 million mf carriers and 19 million chronic filariasis cases.

The study of transmission dynamics of filariasis is complex involving man, vector and parasite and several attempts have been made to study the relationship between host, vector and parasite. However due to multiplicity of variables and processes which are involved makes it difficult to arrive at a comprehensive quantitative understanding of the epidemiology and control of filariasis without the aid of mathematical/ statistical tools.

In the present communication we describe the statistical methods useful for collecting, analyzing and interpreting data for understanding epidemiology of filariasis and its control. The methods that are currently used have been published in the form of manuals and books [1-3,5]. In this manual we confine ourselves to the new mathematical/statistical methods that are developed for analyzing the large data base at the Vector Control Research Centre, Pondicherry. This is expected to meet the needs of not only planners and implementor of national filariasis control programmes but also researchers engaged in filariasis studies. It is hoped to stimulate both

biologists and applied mathematicians/ statisticians to improve our present level of understanding of the dynamics of lymphatic filariasis.

2. COLLECTION OF PARASITOLOGICAL DATA

2.1. THE NEED FOR A SAMPLING DESIGN:

In any epidemiological investigation it is necessary to measure the prevalence or incidence of the disease. Theoretically this could be done by examining all the individuals in a population. Such measurements are generally called *'census'*. A complete survey of all the individuals in a community will be useful not only to detect and treat the carriers but also to have an idea of the extent of the problem under study. As for as the first objective is concerned, it will be fulfilled, provided at least a majority of the population is sampled, if not a 100 percent survey. However 100 percent coverage is impracticable when the number of individuals in the population is very large. The complete survey of individuals for meeting the second objective is of limited value in terms of not only both cost and time but of less precision of the estimates. In such cases a small collection from the larger population will be studied and inference can be drawn about the population. This process is usually called *'sampling'*. A survey following a properly designed probability sampling can provide reliable estimates of disease prevalence or incidence with a greater precision within the available resources.

The conventional blood surveys, in filariasis control programmes, are neither census nor sample survey. These surveys are aimed at collecting as many slides as possible, without resorting to any specific method of data collection. The main objective of mass blood survey is to identify all the parasite carriers in a community and to eliminate the foci of infection either by drug therapy or by isolating the carriers. Using such data for measuring prevalence or incidence of microfilaria in the population may give erroneous picture due to its non-random sampling procedures. These surveys are subject to considerable amount of errors which could be avoided by adopting the scientifically more valid sampling techniques. Further, it does not provide any objective basis for measuring the precision of the sample results. The precision is determined and measured in terms of the standard deviation of the sample estimate. The commonest error is the indiscriminate use of *'pq/n'* (where *'p'* is the proportion of positives, $q = 1\text{-}p$ and *'n'* is the sample size) as the variance of a proportion as a measure of precision, without regard to the manner in which the sample was selected. **If the distribution of disease is very sporadic or concentrated in a small part of the region, a sampling design based on the probability distribution of the disease would be more suitable than a mass blood survey based on stratified random sampling of individuals.**

Census or total survey is neither feasible nor necessary for estimating disease parameters, mainly because of cost and time. The more reliable and scientific estimates of the parameters based on sample survey can considerably reduce the cost and time. However, the adequacy of the sample and reliability of estimates derived from it are guaranteed only if carefully selected probability sampling designs are used. Although many sample designs are available for a particular problem, it is necessary to make a rational choice among them. A rational choice can be made only if comparative efficiency of various sampling designs are known. Eventhough the design of sampling may vary from place to place depending on the objective of the study, as a general guideline the following steps must be followed to get reliable steps[6-8]:

2.1.1. Principal steps involved in a sample survey:

a) Define the objective of the survey:

Since the sampling design and the entire

procedure rests on the objective of the survey it is important to define the purpose of the survey (e.g. Estimate the prevalence/ incidence or detect all parasite carriers etc.).

b) Define the population to be sampled:

The population to be sampled should be clearly defined and it depends on the objectives of the study. For example if the objective is to evaluate the effect of vector control on filariasis transmission, then a longitudinal follow up of the population of children between 0-8 years of age are the population to be sampled. However, if the objective is to *identify all the parasite carriers* for chemotherapeutic control, then a *complete survey* of the population is necessary. If the study is of analytical in nature (eg. studying the prevalence, to monitor and evaluate impact of control programmes in the population) then a cross section of the entire population can be sampled.

c) Data to be collected:

Proper care should be taken to include all the relevant information for the study. Relevant informations are those that may have impact on the study variables either directly or indirectly.

For example: An epidemiological evaluation of the effect of vector control programme for filariasis can be measured by assessing the infection prevalence and intensity of microfilaria before and after control. Since prevalence of infection in a locality depends on many factors such as age and sex distribution of the population, mosquitogenic conditions in the locality, socioeconomic conditions, family size and method of control etc., it is essential to collect information on the above factors. Collecting unwanted and irrelevant information may lead to unnecessary confusion and thus affect the quality of the study.

d) Degree of precision desired:

Since only a part of the population is surveyed in sample study, the results are subjected to some degree of uncertainty. Though it can be reduced considerably by taking large samples and by using more sensitive methods of measurement (eg. membrane filter technique, ELISA techniques for parasitological surveys in filariasis studies), it needs large amount of money and takes more time. Consequently, based on the degree of precision required in the results, sample size will vary. This can be fixed by the researcher in consultation with a statistician.

e) Methods of measurement:

Several methods may be available for measuring the variables under study. The researcher should be well aware of the limitation of the different techniques available for measurement.

For example, a filariasis parasitological survey can be carried out by finger pricking (either 20 cmm or 60 cmm), membrane filter technique and ELISA. Membrane filter technique and ELISA may be more sensitive techniques but in large scale surveys they are less cost effective and time consuming. On the other hand finger pricking is both cost effective and less time consuming, it is not a sensitive method to detect very low microfilaria carriers. Hence a standard method of collecting information is always advisable for any study. A major part of the preliminary work is the construction of record forms on which the questions and answers are to be entered.

f) The frame:

Defining the sampling unit is important before selecting the sample. These are the parts of the population which should not overlap with each other. These may be an individual or household etc. The construction of such list is called sampling frame, is one of the major practical problems in sampling.

Example:In parasitological surveys each individual is a sampling unit. Since construction of a sampling frame in which information about each individual is very difficult to obtain before conducting a survey, each family may be considered as a sampling unit. Hence it is sufficient to have a sampling frame consisting of information about the head of each family.

g) The pretest:

It is always advisable to test the questionnaire in the field on a small scale. This will give an idea about the cost and time required to complete the study. The problems encountered in this can be rectified to a greater extent in the main study and thus reducing the measurement bias.

h) Organization of the field work:

Before starting the survey the field staff should be trained with the objective and with the methods of measurement. Proper supervision of the field staff and their work will guarantee the collected information. Steps should be taken in advance to minimize the non-response bias and to tackle such situations.

i) Summary and analysis of data:

The significant part of any study is the analysis of data. Proper application of statistics for obtaining results is important. Improper application of statistics may lead to wrong conclusion about the population. The following are the sequence of statistical operations to be carried out on any field collected raw data[9]:

(a) Investigation in to validity of assumptions, biases, etc. must be carried out.

(b) Descriptive statistics: The descriptive statistics such as mean, standard deviation, can be presented in tabular form. Graphical representations can be included wherever necessary.

(c) Standard indices: Indices such as infection prevalence, intensity, resting mosquito density, etc. can be computed.

(d) Parameter estimation: Some sensitive parameters can be estimated more accurately with help of mathematical models.

(e) Trends: Trend analysis may be carried out to see the changes in the parameters over a period of time (e.g., change in prevalence of microfilaraemia).

(f) Correlation and Regression: Correlation analysis can be carried out to see the degree of relationship between dependent and independent variables and the regression analysis will be useful to identify the predictor variables and to establish relationship, if any, between the study and predictor variables.

(g) Contigency: Contigency tables can be prepared to analyse the qualitative data using chisquare tests, etc.

j) Information gained for future surveys:

Any completed sample survey is potentially a guide to improve further sampling. The difficulties encountered, the level of error and other information about the population will improve the precision of estimates in the future sampling.

2.1.2. Use of supplementary information:

Supplementary information may be used to provide greater precision in our sample estimates. Stratification, a method for utilizing supplemental information, based on past data may be used to divide the population into groups such that elements in each group are more alike than are the elements in the population as a whole. In case of filariasis control programmes there exist a greater heterogeneity in the transmission potential of different localities in an urban situation. This heterogeneity is the result of unplanned urbanization and hence some areas are more prone to the risk of transmission[10]. Any sampling procedure for filariasis without giving serious thoughts to the following factors would lead to over/under estimation of the parâmeters under study:

a) The distribution of microfilaria is known to be overdispersed in all age groups of the human population.

b) The microfilaria carriers are over crowded due to the focal transmission of the parasite.

c) The conventional blood sampling technique known to provide bias in the estimation of parameters of filariasis control programmes.

2.1.3. Probability sampling methods:

The conventional probability sampling methods are available in many text books[6,7,11-13]. Here we describe those methods with their advantages and disadvantages while conducting surveys. A population can be sampled by using any one or combination of the following probability sampling procedures:.

a) Simple Random Sampling (SRS)
b) Systematic sampling
c) Stratified random sampling
d) Cluster sampling

Simple Random Sampling: This is a method in which each individual in the population has an equal chance of being included in the sample. It needs a list of serially numbered identification of individuals in the population. Individuals are included in the sample by selecting a sample of numbers from random number tables. SRS is simple in theory but is of less practicable when large scale surveys are undertaken. It needs a complete sampling frame of all the individuals in a population.

Disadvantages of Simple Random Sampling (SRS) of individuals:

1. Since in a SRS procedure, individuals are considered as sampling units, the preparation of sampling frame which needs the information about each sampling unit is difficult when large scale surveys are undertaken.

2. In SRS only selected individuals in a house or locality are surveyed. This method, which is restricted only to specific individuals in a house, creates suspicion among the individuals in a family which might affect the precision of the estimate due to the non response bias.

3. Since filarial parasite carriers are known to be aggregated, the SRS of individuals as sampling units fails to give an estimate of aggregation of parasite carriers in the community.

Systematic sampling: In this procedure the selected sampling units are spaced regularly throughout the sampling frame. A starting number 'r' is chosen randomly from a sampling frame and thereafter every r^{th} individual is selected. For example to get a sample of 500 persons from a population of 10000 we proceed as follows:

(a) calculate the random interval,
$k = N/n = 10000/500 = 20$
(b) Select a number 'r' from 1 to k using random number tables.
(c) Select every r^{th} unit from the sampling frame.
(d) Continue step (c) until the required sample size is reached.

Systematic sampling has three major advantages:

i) The investigator does not need to know the sampling frame in advance; the frame can be constructed as the study progresses.

ii) It is simple to implement under field conditions. For example it is easier to take blood smears from every fifth house in a block than to determine which houses are to be visited by means of random number tables.

iii) If a trend is present in the sampling frame, such as from small units to large units, a systematic sample will ensure coverage of the spectrum of units.

For example: A systematic sample of hospital admissions selected over a period of time may provide a better estimate of prevalence of a communicable disease than an estimate based on SRS.

A potential disadvantage of using systematic sample is that it may provide a biased sample if the population contains a periodic type of variation, and if the interval between successive units in the systematic sample happens to coincide with the wave length.

Stratified random sampling: In this procedure the population is divided into strata, or groups of units having certain characteristics in common, and a sample of units is drawn from each stratum. The samples from each stratum can be selected using any one of the probability sampling methods. The method of selection of samples need not be identical in all the strata. For example the investigator can apply the SRS technique in one stratum, cluster sampling in the next stratum and systematic sampling in another stratum etc. Accordingly the sampling procedure is called as Stratified SRS, Stratified cluster and Stratified systematic sampling procedure etc. Stratified sampling is widely used and has several advantages over SRS.

i) The investigator can ensure that each subgroup of the population is represented.

ii) Since the population is divided into homogeneous subgroups, more precise estimates of the population parameters can be obtained.

The main potential disadvantage of stratified sampling is that loss of precision can occur if very small numbers of units are sampled within individual strata.

Cluster sampling: The individuals in population are often grouped into clusters (e.g.,families, villages, hospital wards etc.). In such situations the cluster sampling is more suitable than any other sampling procedures. In this procedure a sample of clusters (e.g.,city blocks) are randomly selected from the population and then observations are made on all individual units (households) in each chosen cluster. Though sampling variance is greater for cluster sampling, it could be reduced by increasing the number of clusters to be sampled. This procedure has two major advantages:

i) Sampling frame is not required for the entire population.

ii) Economic and less time consuming.

Example: Suppose the objective is to study the prevalence of malaria or filariasis in a district. Assume that villages are grouped to form clusters which are identical in all ecological parameters. Then a sample of clusters may be drawn and all the individuals in the chosen clusters are examined.

2.1.4. The sample design for filariasis survey:

Considering the above factors it would be ideal to stratify the study area according to the microfilaria prevalence or entomological observations (zones). In the absence of parasitological and entomological information about the study area, stratification based on sanitary and living condition of the people may be considered. As a second stage, each stratum may be further classified into substrata (eg: streets) consisting of homogeneous sampling units. The sample size for each substratum has to be determined either using prior estimates of the population characteristics if available or in proportion to the population of each substratum. The sampling unit may represent the households which forms a cluster of individuals within the family. Further sampling may be carried out by adopting any one of the following schemes:

Scheme 1.Sampling the substratum:

Each stratum is considered as a cluster of homogeneous streets (substrata). A sample of clusters are selected from each stratum based on simple random sampling without replacement (SRSWOR) technique. Since in this procedure only the substrata are sampled and all the individuals in the selected substratum are surveyed this is called a single stage cluster sampling within the stratum.

Scheme 2.Sampling the households:

In this procedure the sampling is carried out at each substratum level. Each zone (stratum) is further stratified into substrata, consisting of cluster of homogeneous households. Then sample of households for each substratum are selected either by SRSWOR or Systematic sampling methods. The selected households in all substrata

are surveyed and each household is considered as a cluster.

2.1.5. The Estimation of Sample Size:

The method of estimation of sample size varies according to the availability of past information about the population characteristics. As stated earlier if no prior estimates of the population characteristics are available, the sample size for each stratum has to be determined in proportion to the population of that stratum.

If scheme 1 is decided to adopt as the sampling protocol then a sample of substratum is selected. The number of clusters (streets) to be sampled in each stratum is based on the number of households to be surveyed in the zone and the number of houses in the cluster. This procedure is carried out as follows. Based on the average family size, the number of households to be surveyed for a zone is calculated from the zone sample size. Let 'a' be the number of households to be sampled in the zone. Select a street within the zone randomly and note the total number of houses in that street. Continue the sampling of streets until the number of households of the selected streets equals to 'a'. This gives the number of streets (clusters) to be sampled for the zone. Then survey all the households in the selected clusters.

In *scheme* 2, the sample size of the zone is allotted to each substratum according to their population. The number of households are determined based on the average family size of the street. The units (households) are randomly selected using random number tables. Survey is carried out only for the selected houses in all streets.

The principal steps involved in the choice of a sample size are well described in many *textbooks*[6-8,11]. Initially, the sample size is estimated based on the total population of the study area. Then, this is distributed to different strata and substrata according to their population size. Since microfilaria count/carriers are known to be highly aggregated and this distribution is adequately described by the negative binomial probability law, the calculation of sample size must take into account of this factor. The sample size is determined by the following factors:

a) Mean number of mf count/carriers per person/house (m).
b) The negative binomial dispersion parameter (k).
c) Desired level of precision (r) as a percentage.
d) Probability of achieving the desired level of precision (α).

The sample size can be estimated by the following formula:

$$n = (100t_\alpha)^2(1/m + 1/k)/r^2 \quad \text{.....(1)}$$

Where n = Sample size required.
t_α = Students's t value for n-1 degrees of freedom at α % level of significance.
m = Mean number of mf carriers/mf counts
k = $m^2/(s^2-m)$ .....(2)
s^2 = variance of mf carriers/mf counts.
r = desired level of error (%).

For most practical purposes we can assume $t_\alpha = 2.0$ (for 95% confidence limits). Then the formula (1) is simplified to,

$$n = 40000 \times (1/m + 1/k)/r^2 \quad \text{.....(3)}$$

Example: The mean mf count and variance based on the 1981 filariasis survey in Pondicherry are given as 0.9125 and 31.9378 respectively. From this, the value of 'k' was estimated by using equation (2) as 0.0268. Based on these informations the sample size for a filariasis study in Pondicherry can be estimated using equation (3) as 15342.

This sample size can be distributed to the entire study area which was divided into six zones (stratum) with homogeneous sanitary conditions, according to their population size. The number of households to be sampled is determined based on the average family size and the street population accordingly. The households to be surveyed are selected using random number tables and then all the members in the selected houses are surveyed.

There are many situations in which the prior information about the population characteristics

may not be available. In such situations either a pilot study may be carried out to have an idea about the estimate of the parameters or a deliberate selection of about 10% of the population may be sampled.

2.1.6. Parameter Estimation:

Scheme 1:

Let 'm_{ijk}' be the number of individuals in the k^{th} household of the j^{th} street (cluster) in i^{th} zone. Then,

a) the street wise mf prevalence (p_{ij}), which is the proportion of individuals positive for microfilaraemia in the j^{th} cluster unit of i^{th} zone is given by

$$p_{ij} = \sum_{k} a_{ijk} / \sum_{k} m_{ijk} \times 100$$

b) the Zone wise mf prevalence (p_i), which is the ratio of number of mf carriers in the i^{th} zone to total examined in that zone is given by

$$p_i = \sum_{jk} \sum a_{ijk} / \sum_{jk} \sum m_{ijk} \times 100$$

c) Over all mf prevalence is given by

$$p = \sum_{ijk} \sum \sum a_{ijk} / \sum_{ijk} \sum \sum m_{ijk} \times 100$$

$$\text{i.e. } p = \frac{\text{Total no.of +ves in all zones}}{\text{Total number surveyed in all zones}} \times 100$$

Where,
a_{ijk} is the number of mf carriers in the k^{th} household of j^{th} street in i^{th} zone.

Structurally all these are typical ratio estimates which are slightly biased, though the bias is seldom likely to be of practical importance. The estimated variance of the overall prevalence 'p' is given by,

$$v(p) = \frac{(1-f)\{\sum a_i^2 - 2p\sum a_i m_i + p^2 \sum m_i^2\}}{nm^2(n-1)}$$

where,

$m = \sum m_i / n$
$= $ Average number of individuals per cluster.
$a_i = $ Number of mf carriers in the i^{th} zone.
$n = $ No.of clusters and $f = n/N$.

Scheme 2:

Let 'm_{ijk}' be the number of individuals in the k^{th} household (cluster) of j^{th} street in i^{th} zone. Then,

a) the street wise mf prevalence (p_{ij}), which is the proportion of individuals positive for microfilaraemia in the j^{th} street of i^{th} zone is given by

$$p_{ij} = \sum_{k} a_{ijk} / \sum_{k} m_{ijk} \times 100$$

b) the Zone wise mf prevalence (p_i), which is the ratio of number of mf carriers in the i^{th} zone to total examined in that zone is given by

$$p_i = \sum_{jk} \sum a_{ijk} / \sum_{jk} \sum m_{ijk} \times 100$$

c) Over all mf prevalence is given by

$$p = \sum_{ijk} \sum \sum a_{ijk} / \sum_{ijk} \sum \sum m_{ijk} \times 100$$

$$\text{i.e. } p = \frac{\text{Total no.of +ves in all zones}}{\text{Total number surveyed in all zones}} \times 100$$

The estimated variance of the overall prevalence 'p' is as described in scheme 1.

2.2. SAMPLING DESIGN FOR CLINICAL SURVEY

A clinical survey conducted along with blood survey would provide valuable informations relating disease and microfilaraemia positivity. The main purpose of clinical survey is to study the prevalence of disease and to estimate the overall endemicity (includes mf +ves and disease +ves) which is an indication of the extent of the filarial

problem in a community. Recent studies at *VCRC* have used the clinical data for estimating the true incidence by correcting the loss of infection which includes individuals who are apparently negative but clinically positive (*see section - 2.4*). Further the progression of disease in relation to microfilaria status can be evaluated (*see section - 2.5*). The following clinical parameters are useful in the assessment of disease occurrence in a community[2]:

(a) the incidence of acute manifestations such as *adenolymphangitis, epididymo-orchitis, filarial fever, etc.* and

(b) the prevalence of above acute manifestations as well as chronic manifestations such as *lymphoedema (and its grades), hydrocele, chyluria, etc.*

The modified method of Kimura *et al*[14] can be applied for estimating the incidence of acute manifestations[15].

To generate data which are comparable, the clinical survey should take care of the following:

(a) Defined criteria of disease[5,16,17]. All physicians should follow same criteria and should be trained specifically prior to actual survey, if necessary.

(b) Should include both acute and chronic manifestations and actual physical examination should be carried out.

(c) Data of clinical history is useful but should be analysed separately.

Though it would be ideal to examine clinically all persons covered during the blood surveys, it may not be practicable due to the following reasons:

(a) Clinical examination of individuals is time consuming.

(b) Availability of physicians for conducting the survey.

Hence a subsample of blood smeared individuals may be examined clinically and whose history may be recorded. The actual sample size required can be estimated, if prior estimates are available or a pilot survey can be carried out to have an initial estimate of disease rate. If 'P' is the disease prevalence and 'n' the number samples to surveyed and 'r' the percentage of error allowed in 'P'. Then the sample size 'n' can be estimated as follows:

$$n = \frac{40000\ Q}{r^2 P}$$

Here P, Q and r all are measured in percentages. The above formula assumes that the disease carriers are binomially distributed in the population.

However if the distribution is highly aggregated, then the required sample size may be underestimated. In this situation the sample size is estimated by assuming an aggregated probability distribution such as *negative binomial, lognormal, Sichel etc.*

$$n = 40000 \times (1/m + 1/k)/r^2 \quad \ldots\ldots(3)$$

Where,

m = mean number of carriers per house.

k = the negative binomial dispersion parameter.
$= m^2/(s^2-m)$

s^2 = sample variance.

If the distribution is of random in nature then the sample size has to be estimated by using parameter of the Poisson distribution. Since the mean and variance are equal for the Poisson distribution the formula for determining the sample size is very simple and is given as below:

$$n = 40000 / mr^2$$

where 'm' is the mean of the Poisson distribution.

Example:

The frequency distribution of households according to number of diseased in each house, is given in the following table. Estimate the number of households to be sampled for estimating disease prevalence with a relative precision of 20%.

Number of diseased persons	Number of houses
0	509
1	100
2	24
3	12
4	5
5	4
6	2
7	2
8	1
9	1
10	1
Total	**661**

mean number of carriers per house (m)

$= 0.4191$
$s^2 = 1.1862$
$k = 0.2289$

The number of houses to be sampled (n) is obtained using equation (3) as 675.5. Hence a minimum of 676 households can be examined clinically. A subsample of 676 households may be randomly selected (adopting any one of the sampling protocols discussed earlier under parasitological data collection) from the list of households who have parasitologically examined. This way a parallel parasitological and clinical information can be obtained for the same household.

The following indices will be useful:

$$\text{(a) Disease Rate*} = \frac{\text{No. +ve for disease \#}}{\text{Number examined.}} \times 100.$$

\# Confirmed by specific criteria.

* For estimating the total disease rate both acute and chronic manifestations based on clinical findings (not based on history alone) should be included.

(b) Endemicity rate

$$= \frac{\text{No. mf +ve (or) Disease +ve (or) both +ve}}{\text{No. examined both for mf and disease}} \times 100$$

(c) Incidence per 1000 population can also be estimated by modifying the method given in *section 2.4.* for parasite incidence. It is ideal to consider age group 5-15 years for estimating disease incidence.

(d) Further analyses on disease prevalence in relation to age, sex and microfilaria status can also be done, since age and gender crucially influence the disease status[18]. Further this information will enable to study the stability of the disease in a locality and also to observe changes on time scale due to introduction of intervention measures[19].

2.3. FREQUENCY DISTRIBUTION OF MICROFILARIA COUNTS

2.3.1. Introduction:

Certain degree of uncertainty is common in biological experiments partly due to the unavoidable errors in the measurement processes and partly due to inherent variability in the biological processes[20]. The basic problem of Statistics in theory and practice is to estimate the population values based on the samples collected with minimum error. Though several estimates based on deterministic models are available, they fail to take in to account of the random variation in the data. The probability distributions are useful tools to study the inherent variation in the population process which cannot be directly measured. A trend in recent research on parasite ecology has been the recognition of the importance of dispersion patterns to the population dynamics of host-parasite association[21]. Estimates based on the probability distributions are widely used in biology because these estimates are more realistic and gives better understanding of the population dynamics. Statistical models not

only help to improve our understanding of population dynamics, but also provide a powerful tool for the assessment and evaluation of different approaches to control. If used sensibly, they can generate both qualitative and quantitative guidelines for the application purposes. In recent years fitting of suitable mathematical model which best fits the data finds application in most of the areas. The basic assumption in fitting of Statistical distributions is that the variable under study is random in nature with a definite probability of occurrence. The extensive application of frequency distributions in biology is due to its accuracy and its reliability in prediction purposes. The quantification of sampling errors in parameter estimation which are common in biological phenomenon is the main feature of the probabilistic distributions. Fitting a suitable distribution to the data does not convey any meaningful information unless the rationale behind the probability distribution is biologically justified.

2.3.2. Frequency distribution:

The term frequency is defined as the number of occurrence of an event in an experiment. This is a method of summarizing the field collected raw data in the form of table. Such tables can be represented by some known mathematical formulae which is unique for a particular biological or physical phenomena in nature. Usually the frequencies are expressed as the proportion of occurrence to the total number of occurrences. This is also called relative frequency and total of all relative frequencies should be equal to one. In statistics the distribution of these relative frequencies are called the probability distributions.

Distributions in nature are either continuous or discrete. Height, weight, wing length of an insect, instar duration etc are examples of *continuous distributions*. The *normal, exponential, gamma, log-normal etc.* are statistical distributions which are used to describe the continuous type of measurements. The discrete type of data such as number of eggs in an egg raft, microfilaria counts in a slide, number of infected mosquitoes, number of larvae in a dipper sample etc. can be described statistically by *binomial, Poisson and negative binomial* which are *discrete probability distributions*.

Plants and animals in any geographical area are scattered about in one of three spatial patterns: *Uniform, random or aggregated*. Known mathematical frequency distributions can be used as models for describing these spatial patterns[22]. If the frequency distribution of data from a sample fits one of these models, then:

(a) the spatial dispersion of the population can be described in mathematical terms.

(b) errors of population parameters can be estimated.

(c) temporal and spatial changes in density can be compared.

(d) the effect of environmental factors can be assessed.

The important distributions which are most commonly used for describing discrete data are *Binomial, Poisson and Negative binomial.* An initial guess about the form of distribution is obtained by comparing the mean and variance (in other words the dispersion pattern of the data) of the observed frequencies. A simple measure of the degree of dispersion is the variance to mean ratio of parasite numbers per host[23]. The sample values are said to follow *Binomial* distribution if its mean value is significantly greater than the variance. If the mean and variance are approximately equal then the data is supposed to fit Poisson distribution. There are certain situations where the variance is greater than the mean. In such situations the negative binomial distribution fits the data which is to represent the huge variation in the data. This is the most flexible of several possible distributions when the variance is significantly greater than the mean ($\sigma^2 > \mu$).

2.3.3. Negative Binomial Distribution:

Several mathematical distributions have been used as models for describing aggregated or clumped population and the negative binomial distribution is often a suitable model for invertebrate population. The negative binomial is the counterpart of the positive binomial and therefore the probability series of the negative binomial is

given by the expansion of $(q-p)^{-k}$ where $p = \mu/k$ and $q = 1 + p$. The parameters of the distribution are the arithmetic mean m, and the exponent k. The reciprocal of the exponent k (i.e 1/k) is a measure of the excess variance or clumping of the individuals in a population. As 1/k approaches 0 and k approaches infinity, the distribution converges to the Poisson series ($\sigma^2 \rightarrow \mu$) Conversely, if clumping increases, 1/k approaches infinity ($k \rightarrow 0$) and the distribution converges to the logarithmic series. The general term of the *negative binomial distribution* is ,

$$P(x) = \binom{x+k-1}{k-1} (\mu/(\mu+k))^x \ (k/(k+\mu))^k \ .. \ (1)$$

where, P(x) is the probability of observing x counts in a sampling unit. The k and m are the parameters to be estimated. Negative binomial can be applied to a wide variety of contagious distributions such as,

a) True Contagion: The presence of one individual or event increases the chance that another will occur in the same time.

b) *Constant birth-death-immigration rates:* The growth of a population with constant rates of birth and death per individual and of immigration per unit of time leads to a negative binomial for population size.

c) *Randomly distributed clumps:* If clumps of individuals are distributed at *random* and the number of individuals in the clumps are distributed independently in a *logarithmic distribution,* then a negative binomial will result.

d) *Heterogeneous Poisson sampling:* If the population distribution is of continuous type and if the process of sampling is Poisson, then the resulting distribution is a *compound Poisson or negative binomial* distribution.

2.3.4. Method of Fitting Negative Binomial Distribution:

The fitting of negative binomial distribution involves estimation of the parameters μ and k where,

μ = the mean of the distribution and
k = the dispersion parameter

The parameter m is estimated by the mean '*m*' of the sampling distribution and an approximate value of '*k*' can be obtained by the following formula:

$$k = m^2/(s^2-m) \ .....(2)$$

where m and s^2 are the mean and variance of the observed sampling distribution. An accurate estimate of *k* can be obtained by substituting the value of '*k*' and '*m*' in the *maximum likelihood equation*.

$$n\ln(1+m/k) = \Sigma \ (A_{(x)}/(k+x)) \ .... \ ..(3)$$

where $A_{(x)}$ is the total number of counts exceeding x and n is the total number of sampling units. The probability values for various counts are calculated by substituting the values of m and k in equation (1).

EXAMPLE:

The following are the mf counts distributed in human population collected from an area in Pondicherry.

mf counts	*No. of individuals*
0	300
1	95
2	70
3	50
4	15
5	7
6	4
7	4
8	3
9	2
10	1
TOTAL	551

In this example,

1) since the distribution of mf within human is random, collecting blood smears from different individuals form a mixture of Poisson process which leads to some aggregated distribution.

2) Mean is less than the Variance.
mean = 1.0599 < variance = 2.5681

Hence, the observed frequency distribution may be adequately described by the negative Binomial distribution which is the suitable distribution for representing the large amount of variation in data.

The approximate value of k is obtained from equation (2).

i.e k = 1.0599 ^ 2/(2.5681-1.0599)
= 0.7425.

However, a real estimate of 'k' can be obtained by trial & error method. i.e. by taking the values of k from 0.5 to 1.0 with increment 0.05. For each value of k the LHS and RHS of equation (3) can be estimated. The difference between LHS and RHS can be obtained and tabulated as below.

k	m	LHS	RHS	DIFFERENCE
0.5	1.0599	625.91	660.69	-33.79
0.6	1.0599	560.69	568.73	- 8.05
0.6455	1.0599	535.31	535.77	- 0.46
0.649	1.0599	533.46	533.42	0.04
0.8	1.0599	464.86	450.08	14.78

LHS - Left Hand Side
RHS - Right Hand Side

Hence the value of k at which the sign of difference changes first time can be taken as the real value of 'k'.
Then the estimated m and k values are substituted in the equation (1) to get the probability values for different counts. The expected frequencies are calculated by multiplying the probability value with the total frequencies. A test of significance based on chi square distribution can be obtained by comparing the observed and expected frequencies.

2.3.5. Frequency Distribution of Parasite in Human host:

Helminth life cycles are often complex, involving more than one host species and many distinct parasite development stages. The distribution of parasite numbers per host is important for understanding the transmission dynamics of the parasite and also in the implementation of control policies[24,25]. The frequency distribution of parasite numbers per host have generated considerable interest in helminthology[23]. The majority of distributions have been found to be aggregated or over dispersed and ecological theory strongly suggests that this serves to enhance the density dependant regulations of parasite abundance[21]. Empirical evidences suggest that often the negative binomial probability distribution has proved to be a good descriptor of observed pattern of parasite aggregation[26]. Although, lognormal probability distribution can be used to model the observed pattern, it is less flexible to further mathematical treatment. Frequency distributions are primarily of value in the elucidation of qualitative as opposed to quantitative concepts or principles.

In filariasis, the statistical distribution of worm numbers per host are invariably highly aggregated, with majority of worms often being harbored by a minority of the host population[23]. The distribution pattern of parasite is similar in both definitive (man) and indefinite host (vector) of invertebrate species. The generative mechanisms of such patterns are many and varied, but include genetic, spatial, and behavioral factors. Recently, park (1988) had pointed out that the frequency distribution of mf counts in samples can be described statistically by a compound Poisson distribution (the negative binomial distribution) which reflects Poisson sampling and an underlying (gamma) distribution of mf density in the blood of the human population[27]. However, it was observed that the distribution of mf in

human does not fit the negative binomial distribution. This discrepancy may be attributed to the following reasons:

i) Observed negatives are not true negatives as proposed by Park et.al due to sampling.

ii) The observed negatives include,

a) true negatives who are completely without macrofilarial infection and

b) single sex and prepatent infections.

Grenfell et al (1989) have refined this analysis, proposing the use of a zero truncated negative binomial distribution to estimate the considerable proportion of hosts who appear to have no mf in their blood[28]. This modification will therefore affect estimates of the prevalence of infection. Das *et al* have applied this refined analysis and estimated the excess zeros which are due to sampling bias using the truncated negative binomial distribution[29].

2.3.6. Estimation of Excess Zeros in the sample:

The observed mf frequency distribution for each age group is used to fit a zero truncated negative binomial distribution by maximum likelihood method. The estimated zero probability of this distribution, p'(0) can then be used to calculate Z, the proportion of the age group who have a true zero count (in excess of the proportion due to blood sampling), from the equation,

$$Z = [p(0) - p'(0)] / [1 - p'(0)] \quad .....(4)$$

Where p(0) is the actual proportion of negative samples. Finally we can estimate the prevalence (R) of infection, corrected for sampling zeros by the equation

$$R = 1 - Z$$

Example:

The frequency distribution of mf in 15-19 year age class of the people in Pondicherry was fitted to the zero truncated negative binomial distribution and the estimates of parameters μ and k were found to be 6.26 and 0.27 respectively. The proportion of observed zeros (negatives) was 0.895. The estimated proportion of zeros (p'(0)) is the probability of observing zero counts using truncated negative binomial distribution. These values are substituted in equation (4) to get the excess zeros 'Z' due to sampling as

i.e $Z = (0.895 - 0.426)/(1-0.426) = 0.817$

The corrected prevalence = 1-0.817 = 0.183. Similar estimation can be made for all age groups.

2.4. ESTIMATION OF AGE SPECIFIC RATES OF ACQUISITION AND LOSS OF INFECTION.

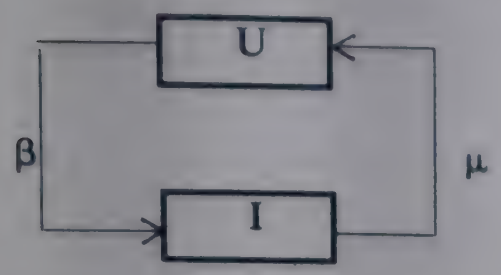
2.4.1. Introduction:

In bancroftian filariasis, the age-distribution of infection is remarkably consistent throughout a wide geographical range[3]. Typically, both prevalence, and intensity rise steadily to a maximum in the 15-25 year age class and then decline slightly to a stable plateau in adulthood. Such patterns appear to be generated by a simple immigration-death process[30] in which the *Rate of gain of infection* (determined by such factors as vector biting rate and host resistance) is counteracted by the *Rate of loss of infection* (perhaps determined by the host's ability to terminate infection). Rates of acquisition and loss are sensitive parameter not only for understanding dynamics of parasite population but also for evaluating control programme. Such age prevalence data can be analyzed to estimate the rates of acquiring and losing the infection by appropriate *Catalytic Model* developed by Muench [31].

2.4.2. Catalytic Models:

This type of models were originally described by Muench[31] to describe certain chemical reactions and later on applied to describe dynamics of infectious diseases. The basic concepts of Muench's catalytic model is described below. In an endemic area for any disease the equilibrium dynamics of uninfected (U) and infected (I) human popula-

tion may be described as follows:



Mathematically the status of infected population (I_t) over a time period 't' can be expressed by the following formula

I_t = Initial infected (I_0) + persons gained the infection during time period 't' - Persons lost the infection during time t

Changes in the infected population over a time 't' can be described as

$I_t - I_0$ = Persons gained infection - persons lost the infection.

Thus Change in the infected population per unit time can be calculated by dividing with total time period 't'.

$(I_t - I_0)/t$ = (persons gained infection -persons lost the infection)/ t

In terms of differential equation it can be expressed as

$$dI/dt = \beta U - \mu I \qquad (1)$$

Similarly the status of uninfected population (U_t) after a time period 't' will be

U_t = Initial uninfected population (U_0) + persons lost the infection during time 't' - persons gained the infection over a time period 't'

Changes in uninfected population can be described as

$U_t - U_0$ = persons lost the infection - persons gained the infection

Thus changes in the uninfected population per unit time will be

$(U_t - U_0)/t$ = (persons lost the infection - persons gained the infection)/t

The changes in the uninfected population can be expressed in the form of differential equation as

$$dU/dt = -\beta U + \mu I \qquad (2)$$

where,
μ is the instantaneous rate of loss of infection from an infected individual;

and β is the instantaneous rate of gain of infection by an uninfected individual.

Thus changes in the infection status of a host population can be calculated from the rates of loss and gain of infection.

The assumptions of Muench's models:

The use of Muench's models depend upon several assumptions.

a) the most fundamental assumption is that the epidemiological situation has remained constant over the time period t.

b) the rate of acquiring the infection is same for whole population.

c) the rate of acquisition and loss of infection is constant.

d) the rate of acquisition is the same for those persons who have lost the infection as for those who never had it or the prior infection does not alter the rate at which negative people become positive.

Based on the above assumptions three basic models were proposed.

(a) **simple catalytic model,(one way reaction)**

(b) reversible catalytic model,(two way reaction)

(c) two stage catalytic model (successive reaction)

The above three basic models are discussed below with examples.

(a) In the simple catalytic model,the proportion of persons showing the sign of infection can be expressed by the equation:

$$y = 1\text{-}e^{-\beta t} \text{ ------(3)}$$

where y is the proportion of persons showing the sign of infection after a time period 't'years in an area where the rate of acquisition per person per year is β. In this case, it is assumed that the people in the endemic area are exposed to the risk of infection at the same rate, and the method of diagnosis detects all infected persons.

Eg. Viral infections, where detection of all cases is possible.

This model considers the rate of acquisition and not the rate of loss, therefore this model is useful in describing such disease where recovery is not at all possible like Diabetics and Hypertension.

(b) In the reversible catalytic model, it is assumed that the persons in an endemic area are exposed to a constant risk of infection, conversion of negatives to positives taking place at a constant rate (β), and conversion of positives to negatives taking place at a constant rate of 'μ'. Then the proportion of persons (y) showing infection after a time period of 't' is expressed by the equation:

$$y = \{\beta/(\beta + \mu)\}\{1\text{-}e^{-(\beta + \mu)t}\} \text{ -------(4)}$$

Eg. This model can be applied for diseases like malaria and filariasis where conversion of negatives to positives and positives to negatives takes place at a constant rates.

(c) In the two-stage catalytic model, it is assumed that persons living in an area endemic for infectious disease are exposed to a constant risk of infection, conversion of negatives into positives takes place at a constant rate of β, positives continue to lose the infection at a rate of μ and those who lose infections are not susceptible to reinfection again.

Eg. Persons infected with smallpox can become uninfected and they will never become infected again for the same disease.

In this case, the proportion of persons (y) over a time period (t) showing the positive sign of the infection is expressed by the equation:

$$y = \{\beta/(\beta - \mu)\}(e^{-\mu t}\text{-}e^{-\beta t}) \text{ ------(5)}$$

It is obvious that all of these mathematical models are constructed on simple hypotheses neglecting the effects of various complicated biological factors which differ from disease to disease (like viral infection, bacterial, protozoal and helminthic diseases). Therefore appropriate model need to be selected for studying population dynamics of pathogen.

2.4.3. Application of Catalytic Models in Lymphatic Filariasis:

Simple catalytic model does not take into account the loss of infection, whereas in lymphatic filariasis loss and gain of infection is a common feature. Secondly, in the case of lymphatic filariasis 100% of the infected people can not be detected by the present techniques. Hence this model can not be used for describing dynamics of lymphatic filariasis. Similarly two stage model assumes that the reinfection do not occur whereas in lymphatic filariasis reinfection is a common feature. Therefore this model also can not be used. On the other hand reversible catalytic model appears to be more appropriate in describing the dynamics of *W.bancrofti* because this model assumes the chance of reinfection in recovered people. Based on the above reasonings the reversible catalytic model can be used for the estimation of the rates of acquisition and loss of *W.bancrofti* infection[32,33].

For age specific rate of acquisition paired sample should be collected from an area over a period of time. Ideally the area should have reached a stable state and acquisition of infection is interrupted by control measures. The data in the following example are obtained from a control area where transmission is interrupted by efficient vector control measures[34].

Example:

Data source: control area.

Age (yrs)	Sampled (81 & 86)	Positive in 81	No.became negative in 5 yrs
0-5	374	10	6
6-10	950	55	30
11-15	796	83	52
16-20	537	70	44
21-25	400	56	32
26-30	426	43	25
31-40	633	50	36
41-50	451	35	19
> =51	310	24	15
Total	4877	426	259

As the rate of acquisition is nil (i.e. $\beta = 0$) in the control area then the equation $dI/dt = \beta U - \mu I$ will become $dI/dt = -\mu I$. By solving the equation the rate of loss (μ) can be calculated by the following formula:

$$\mu = -\ln(I_t/I_0)/t \text{ ---------(6)}$$

where 'ln' refers to the natural logarithm and I_t/I_0 refers to the proportion still infected (i.e. from 1981 to 1986) up to time period 't' (in this case t = 5years).

$$\frac{\text{No. of persons +ve both in 1981 \& 1986}}{\text{No. of +ve persons in 1981}}$$

Considering the total cases, out of 426 mf +ve persons in 1981 167 remained mf positive in 1986 also.

Proportion still infected for 5 years

$$= 167/426$$
$$= 0.392$$

Instantaneous rate of loss of infection (i.e. μ) is

$$\mu = -\ln(0.392)/5$$
$$= -(-0.9365)/5$$
$$= 0.1873$$

2.4.4. Calculation of the life-span of adult worm:

A more straight forward approach for estimating the life-span of adult worm has involved long term observation of infected patients no longer exposed to reinfection either because they have moved out of the endemic area or because local transmission has been completely stopped. The period during which microfilariae were detectable in blood can be calculated by taking the reciprocal of μ (i.e. $1/\mu$)

The life span of adult worm ($1/\mu$)

$$= 1/0.1873$$
$$= 5.34 \text{ years}$$

2.4.5. Calculation of gain of infection:

For estimation of the age specific rate of gain of infection paired data can be collected from individuals living in an endemic area where no control measures are undertaken. Information about the status of parasitaemia in each individual has to be obtained at two different time points with an interval of not less than the prepatent period of the parasite under consideration.

From the following data the gain of infection can be estimated as follows:

$$\text{The gain of infection} = \frac{\text{No. of persons -ve in 1981 became +ve in 1986}}{\text{No. of -ve persons in 1981}}$$

Data source (whole Pondicherry endemic area)

Age(yrs)	Sample	mf-ve 1981	-ve ('81) to +ve ('86)
0-5	507	494	15
6-10	1404	1329	42
11-15	1207	1084	45
16-20	824	730	38
21-25	616	535	21
26-30	671	606	20
31-40	994	914	24
41-50	745	685	23
>50	557	517	18
Total	7525	6894	246

In the Endemic Area, about 0.036 (246/6894) of the amicrofilaraemic cohort in 1981 had become microfilaria positive by 1986. Now, the change in uninfected population over a time 't' is given by the equation (2) and the solution of this equation (ie. proportion of uninfected population over time 't') is U_t/U_0, i.e .

$$U_t/U_0 = e^{-(\beta+\mu)t} + \{\mu/(\beta+\mu)\}\{1-e-(\beta+\mu)t\} \quad \ldots\ldots\ldots\ldots(7)$$

by subtracting U_t/U_0 from 1 we can get the proportion of persons showing infection after a time period 't'. ie.
$1-U_t/U_0$.

An estimate of the rate of acquisition of infection for an equivalent population in which transmission continues can then be obtained by substituting this estimate of μ in the equation (7). However, the real estimate of rate of acquisition (β) can be obtained by trial & error method i.e. by taking the values of β from 0.008 to 2.0 with increment 0.0002. For each value of β, U_t/U_o can be estimated from equation (7). The difference between the proportion of observed gain (0.036) and estimated gain ($1-U_t/U_o$) can be obtained and tabulated in the following manner.

The value of β at which the sign of difference changes first time can be taken as the real rate of acquisition.

Estimated β	U_t/U_o	Difference
0.008	0.97	0.010017
0.0082	0.97	0.009386
-	-	-
-	-	-
0.0112	0.96	-0.00001
0.0114	0.96	-0.00063

In practice, however, the apparently uninfected and presumed susceptible, population contains a proportion (x) of individuals who have a history of infection but have become microfilaria negative during the progression to disease. These individuals are potentially resistant to further infection and may introduce bias into the estimation of the size of the susceptible populations. This potential source of bias may be discounted by assuming that disease rate in amicrofilaraemic individuals is constant over 5 years period.

The proportion of actual susceptible population over a time period 't' is given by

$$S_t/S_0 = (U_t - x_t)/(U_0 - x_0)$$

Where S_t is susceptible population at time t.
S_0 is initial susceptible population
U_t is uninfected population at time t
U_0 is uninfected population at initial time
x_t is number of clinical sign cases among uninfected population at time t
x_0 is number of clinical sign cases among uninfected population at initial time.

The proportion of amicrofilaraemic individuals with chronic disease in each age-class should be determined by the clinical and blood examination of a random, age-stratified sample.

2.4.6 Calculation of corrected gain:

Assume that disease pattern is same from 1981 to 1986. The actual amicrofilaraemic persons (S_o) in 1981 is given by

No. of -ves in 1981 (U_o) - (U_o x proportion of disease in -ves (x_o)).

Data source for corrected rate of acquisition:

Endemic area (males)				Clinical survey (males)	
Age (Yrs)	Sample	mf-ve in '81	-ve ('81) +ve ('86)	Sample	prop mf-ve with clinical signs.
0-5	251	245	9	114	0.00
6-10	700	658	20	228	0.01
11-15	587	520	20	216	0.05
16-20	377	326	22	135	0.11
21-25	245	208	10	86	0.19
26-30	225	193	9	56	0.24
31-50	607	551	21	176	0.33
>50	302	277	7	94	0.43
Total	3294	2978	118	1105	0.14

considering the total cases

ie 2978-(2978 x 0.14) = 2561.08

The actual amicrofilaraemic persons (S_t) in 5 years time

(i.e. in 1986) is given by

(no.of mf-ves in '81- no.with neg. to pos. in 5 years)-(mf-ves(in '81) x prop.dis.in mf-ves)

ie. (2978-118)-(2978 x 0.14) = 2443.08

So, the corrected gain is given by

$$1 - \frac{\text{(actual amicrofilaraemic in '86}(S_t)}{\text{actual amicrofilaraemic in '81}(S_0))}$$

ie 1-(2443.08/2561.08) = 0.04607

For males the loss rate of infection (m) is found to be 0.17746. Having the corrected gain and loss rate of infection for males the corrected rate of acquisition can be obtained by iteration using equation (7) as done for the estimation of β (uncorrected for disease). The corrected rate of acquisition is found to be 0.0144. The estimated mf prevalence is given by

$$\frac{\text{Rate of acquisition x 100}}{\text{loss rate of infection}}$$

This is called point infection prevalence and this can be compared with the observed infection prevalence.

2.5. ESTIMATION OF RISK OF DEVELOPING DISEASE IN RELATION TO MICROFILARAEMIA STATUS, AND, TRUE INFECTION PREVALENCE IN A POPULATION.

In lymphatic filariasis it was observed that most of the people with chronic manifestations are microfilaria (mF) negative on blood examination. Since individuals who had ever been infected are the potential source of developing the disease, establishing a relationship if any, between the infection status and the consequent progression of disease in an infected individual, may provide an estimate of the risk of an infected individual developing the clinical manifestations[35]. The main aim of this analysis is to estimate the risk of acquiring disease in persons who had been

2.5.1. Estimation of the proportion ever infected using catalytic model:

In the reversible catalytic model,

$$y = [\beta/(\beta + \mu)]\,[1 - e^{-(\beta+\mu)t}] \quad \text{.....(1)}$$

assuming that there is no loss of infection i.e., $\mu = 0$, the proportion of population that would have remained mf positive for ever without losing the infection, (c_t) over a period of time 't' is given by

$$c_t = 1 - e^{-\beta t} \quad \text{.....(2)}$$

where 'β' is the instantaneous rate of acquisition of infection. Since the gain rate is age dependent (see section - 2.4), the cumulative infection prevalence has to be estimated for each age class independently. This can be obtained by correcting the proportion who had infected in the previous years of their life. Thus, if u_{tn} is the proportion uninfected at time tn, then the proportion ever infected at time tn + 1 is equal to,

$$c_{tn+1} = 1 - u_{tn} e^{-\beta_{tn+1}(t_{n+1} - t_n)} \quad \text{.....(3)}$$

Where, $u_{tn} = 1 - c_{tn}$.

c_{tn} = proportion ever been mf positive at time 'tn'

β_{tn} = rate of acquisition of infection between the time periods 'tn-1' and 'tn'

Here the two time periods are the mid points of the age classes.

2.5.2. Estimation of the risk of developing clinical symptoms:

Using the age specific rates of loss and acquisition of infection ('μ_t' and 'β_t'), the infection prevalence (corrected for disease) for age 't' is calculated as follows:

$$p_t = \beta_t/\mu_t \quad \text{.....(4)}$$

The proportion of the people who are at the risk of developing the clinical symptoms of filariasis (R_t) can be obtained by subtracting the corrected prevalence from c_t.

i.e. $$R_t = c_t - p_t \quad \text{.....(5)}$$

This proportion 'R_t' of people who had been microfilaria positive earlier and become apparently negative now, (who appear to have recovered from infection) is compared with the chronic disease prevalence. It was seen that in Pondicherry particularly in the male population, the recovery (risk) rate resembled closely the chronic disease prevalence suggesting the possibility that the recovery process triggers off inflammatory reaction and disease.

This corroborates with the immunological theory of lymphatic filariasis which says the response of the immune system to the infection results in the clearance of mF and death of adult worms thereby creating blockades in the lymphatic system which in turn leads to clinical manifestations[36]. The inference made from this analysis is that the people who have become mF negative from being positive are more at risk of developing disease than the ones who are microfilaraemic.

Example :

The reinfection data was used to estimate the age specific rates of acquisition (β) and loss (μ) of infection. The methods of estimation of these two parameters are explained in the previous chapter.

The age specific rates of acquisition and loss of infection for males in the first few age classes are given below:

Age class (in years)	Rate of Loss	Rate of Acquisition
0 - 5	0.0575	0.0087
6 - 10	0.1685	0.0091
11 - 15	0.1788	0.0120
-	-	-
-	-	-

The proportion who have ever been infected in the first age class 0-5, can obtained using equation (3) as follows:

here, to = 0, t1 = 2.5, β_{t1} = 0.0087 and u_{to} = 1

$$c_{2.5} = 1 - e^{(-0.0087 \times 2.5)}$$
$$= 0.0215$$

For 6-10 age class,

here, t1 = 2.5, t2 = 8.0, β_{t2} = 0.0091
and u_{t1} = 1-0.0215

$$c_8 = 1-[0.9785 \times e^{-0.0091 \times 8-2.5}]$$
$$= 1-0.9307$$
$$= 0.0693$$

Proceeding this way, we can calculate the proportion ever positive for all the age classes.

From the proportion ever positive for mF, we next estimate the proportion of the people who are at the risk of developing the clinical symptoms of filariasis. Using the age specific rates of loss and gain of infection ('β_t' and 'μ_t'), the prevalence of infection for the age class 6-10 can be obtained using equation (4).

i.e. $p_8 = 0.0091/0.1685$
$= 0.0540$

The proportion of ever infected people who are at the risk of developing the clinical symptoms of filariasis (R_t) can be calculated using equation (5).

i.e. $R_8 = 0.0693 - 0.0540$
$= 0.0153$

Similarly the estimate of risk of developing the clinical symptoms of filariasis can be obtained for all age classes.

2.5.3. Estimation of true infection prevalence using the probability of acquiring a male or female worm:

The transmission dynamics of filariasis involves three factors, namely, (i) the human host, (ii) the parasite and (iii) the vector. The human population in an endemic area can be either infected or uninfected. The infected population are further subdivided into the following:

a) microfilaraemic,

b) individuals with very low worm burden,

c) single sex (worm) infections

d) individuals at prepatent period and

e) diseased

For any disease, one of the measures of transmission is its infection level in the population. Infection level is defined as the proportion of persons harboring the parasite causing the disease in a population. Conventional method of estimating the disease\infection prevalence identifies only the microfilaraemic population and ignores all other groups (from (c) to (e)) due to following technical difficulties.

In Lymphatic filariasis, a person is said to be infected the moment the infective larvae enters his body. The conventional blood sampling, can not detect the infected persons during the prepatent period. It is also less sensitive to detect carriers with *ultra low level mf* in their blood. Although membrane filter concentration technique facilitated the detection of low density carriers of microfilariae, its application in large scale survey is impracticable due to various factors[3]. Since the parasite is of dioecious in nature, these techniques can not detect persons with single sex (worm) infection. Therefore, it is practically not possible to estimate all the '*infected*' persons in a population for the above mentioned reasons. However, an estimation of infection prevalence which takes account of few of these factors will be useful to define the public health importance of filariasis in an area.

In section 2.3, it has been shown that the fitting of truncated negative binomial model to the frequency distribution of various mf counts in the population also accommodates the single sex and single worm infections while correcting the observed prevalence of infection[28]. In this analysis the true infection prevalence (including single sex (worm) infections) in the population is estimated using the probabilities of occurrence of male or female worms based on some assumptions. The

method uses the age specific rates of loss and gain of infection calculated using the reversible catalytic model (described in detail in section-2.4). Consequently this analysis gives an account of the changes in the mF prevalence due to single sex (worm) and its significance in disease development (as shown by Hairston and Jachowski in the case of aperiodic bancroftian filariasis[32]). The basic assumptions for calculation of probability of acquiring a male or female worm are:

(i) The event of acquisition of a male worm by an individual is independent of him\her acquiring a female worm.

(ii) Any infected individual will harbor equal number of male and female worms i.e., he/she will harbor equal ratio of male and female worms.

Violation of these assumptions might lead to erroneous results and there by making the interpretations difficult.

Let 'p' be the proportion of persons positive for infection as obtained from the microfilaria survey. This proportion of people harbor both male and female worms which is evident from the presence of mf. Based on the assumption (i),the proportion positive for infection should be equal to the product of the proportion harboring male worm (p_m) and proportion harboring female worm (p_f).(Since p_m and p_f are independent). i.e.,

$$p = p_m \times p_f$$

By assumption (ii),

$$p_m = p_f = p_c$$

Therefore,

$$p = p_c^2$$

where p_c is the constant proportion containing either male worms or female worms. Therefore it follows that

$$p_m = p_f = \text{sqrt}(p) = p_c$$

Based on the same arguments, substitute *the proportion ever positive for mf* (c_t) in the place of *proportion positive* (p). We then calculate the proportion who were ever infected with either male worm (y_m) or female worms (y_f) by taking the square root of the proportion who had ever been positive for mf over a period of time 't' i.e.,

$$y_m = y_f = \text{sqrt}(c_t) \quad \ldots\ldots(6)$$

Then the proportion of persons never infected with female worms is given by

$$(1-y_f)$$

Since there is an equal sex ratio among the worms, the proportion which escaped the infections due to male worms also is given by

$$(1-y_f)^2 = y_n \quad \ldots\ldots(7)$$

where y_n is the proportion which escaped all the infections (due to both male and female worms) i.e., the one which is uninfected. The proportion ever infected with worms of either sex, 'y_t' can be estimated by the following expression

$$y_t = 1 - y_n \quad \ldots\ldots(8)$$

Now this 'true' infection prevalence can be calculated for different age classes and is compared with the disease prevalence. It was found that in Pondicherry there was no resemblance between the two suggesting the negligible role of single sex (worm) infections in disease development in the case of nocturnally periodic bancroftian filariasis. However interpretations based on the estimates of true infection should be done with enough care since all these calculations are based on the fact that the assumptions (i) and (ii) are true under all situations.

It should be also noted that even this estimate of true infection prevalence does not include the persons in prepatent period. Therefore to include those individuals also, it is necessary to measure the number of L_3 larvae that successfully enter the human system thorough immunological methods.

Example : The proportion ever infected (c_t) is 0.0215 for the age class 0-5. The true infection

prevalence can be estimated as follows:

Using equation (6), the proportion ever infected with female worms, mated or not is given by

$$y_f = \text{sqrt}(0.0215)$$
$$= 0.1467$$

Then the proportion of persons never infected with female worms is given by,

$$= 1 - 0.1467$$
$$= 0.8533$$

By equation (7), the proportion which escaped the infections of both male and female worms is obtained as below

$$= (0.8533)^2$$
$$= 0.7281$$

Then the true infection prevalence (persons ever infected with worms of either sex, mated or not) is estimated using equation (8),

$$= 1 - 0.7281$$
$$= 0.2719$$

Proceeding in a similar manner, the true infection prevalence can be estimated for all age classes.

2.6. SPATIAL DISTRIBUTION OF PARASITE CARRIERS

In section-2.3 the negative binomial probability distribution was used to describe the distribution of worm aggregation in the human host. Since clustering of infected individuals in households is potentially an important consideration, it is necessary to study the spatial distribution of mf carriers in relation to size of households. Several published reports are available in this aspect[14,37-39]. Ottesen *et al*.. .,(1981) have also reported that the familial clustering is also due to genetic influences on the susceptibility to bancroftian filariasis[37]. Kimura et al ., demonstrated the clustering of mf carriers by considering only those who have been examined (irrespective of family size) for microfilaraemia. Since in their study, all the members were not examined, the proportion infected in each household is biased to the number examined[14]. Vanamail *et al* ., has extended this analysis for each household size, thus any bias resulting from differing proportions of infected individuals in household of different sizes was avoided[39]. In this section the method of collecting and analysing data on the spatial distribution of mf carriers are discussed.

METHODS:-

In an endemic area the house wise enumeration (age & sex) of persons should be carried out prior to any parasitological survey. During enumeration time it is also important to identify the various factors that are supposed to influence the transmission dynamics of the disease with respect to each house.

Format: Enumeration Form for Microfilaria Survey

House No.	Age/ Sex	Breeding Source	History of cases	Other factors

After completing the enumeration with all relevant informations, the parasitological survey should be carried out. During this study as much as possible all persons in the target house should be covered. Equal weightage should be given for all size of households.

Tabulation: From the enumeration and parasitological survey data the following Table should be completed.

Family Size	No. of houses Enumerated	Full y covered	Partly covered
1			
2	(* Example)		
3			
4			
5*	250	117	70

2.6.1. Analysis:

For clustering analysis it is more reliable to consider the data on full participated houses in the study. The results of parasitological survey from the full participated houses can be obtained for each family size as follows.

Family Size	Houses with +ves 0	1	2	3	4
1					
2					
3					
4					
5*	82	30	4	1	0
6					
7					
8					
9					

* - example

For each family size the frequency distribution of parasite carriers can be obtained.

Example: **Frequency distribution of parasite carriers for family size 5.**

No. of infected persons	No. of houses
0	82
1	30
2	4
3	1

In order to see the distribution Pattern of infected persons in relation to family size, statistical distributions can be fitted to the observed data[39]. If the distribution of infected individuals is random then Poisson probability model will describe the observed trend in a significant manner. The form of Poisson distribution is

$$P(x) = e^{-\mu} \mu^x/x! \quad x = 0,1,2,.....$$

Where, 'μ' is the mean of Poisson distribution

If the Poisson distribution do not describe the trend adequately then the distribution of infected persons deviate from randomness and it implies that the distribution of infected individuals is aggregated or clumped. Since negative binomial distribution is one of the clumping distributions, aggregation phenomenon can be tested by fitting negative binomial probability statistics. The method of estimation of parameters of this distribution are described in section 2.3. The testing of randomness or clustering can be carried out for each family size.

Example: **Fitting of Poisson distribution to the above frequency data for family size 5.**

Infected	Observed	Expected	Chi-sqr
0	82	82.4	0.00207
1	30	28.9	0.04344
2	4	5.1	0.08596
3	1	0.6	
Total	117	117.0	0.13147

Here calculated Chi-square value is 0.13147. While comparing this with the tabulated value for 1 degree of freedom (i.e number of observations in the Chi-square column-1-number of parameters to be estimated) the level of goodness is more than 0.05 (ie $P > 0.05$). The goodness of fit indicates that the distribution of infected individuals is random. In certain cases the family sizes can be added into 3 or more groups due to less number of observations in each family size. Adopting the above two statistical approaches one can get two types of households i.e. the households under the family size at which randomness appear and the households under the family size at which clustering phenomenon occurs. The significant causes for difference in distribution pattern can be analyzed further by considering age and gender distribution of the population in the two types of households and also other possible factors that influence the transmission dynamics of the disease.

3. ENTOMOLOGICAL DATA

3.1. Introduction:

Transmission dynamics of any disease involves the study of interaction between the parasite with man and the vector. In the previous chapters we have described the methods of sampling the human population for studying the distribution of parasite in the host and their practical implications in estimating the prevalence and the estimation of fecundic life span and age specific incidence of the parasite in the human host. Further the relationship of microfilaria infection with disease prevalence are explored. In this chapter the methods of sampling the larval and adult stage of the vector population (the intermediate host) and the parasite distribution in the vector, and its relationship with parasite distribution in human population are described.

The various methods of sampling and estimation of the larval and adult population are well described in many text books[11,40,41]. A recent WHO publication has well documented the various methods of sampling and estimating the adult population of *C.quinquefasciatus* [42]. We begin this chapter with the methods of sampling the intermediate host of the parasite. The life cycle of the intermediate host (*C. quinquefasciatus*) involves two stages, namely, the immature stage and adult stage. Since the immature and adult stages are living in entirely different habitats, the sampling methods also varies accordingly.

3.2. SAMPLING THE LARVAL POPULATION:

The purpose of larval sampling is of many fold. The following are the few among many objectives:

1. To monitor and give feed back to control operations.

2. To study the dynamics of the larval population.

3. To study the relative importance of species composition in the transmission of the disease[43].

4. To study the relative importance of different types of habitats for breeding of a particular species[43].

The choice between one time survey or continuous surveillance of the immature population will depend on the objectives[42]. If the objective is to give feed back to larval control operations then a continuous surveillance is necessary. The second and third objectives may be fulfilled by one time or continuous surveillance of the larval and adult population. An account of the need for a sampling design and the advantage of using stratified random sampling are described with suitable examples in " *A note on Sampling methods* " by S.J.Poti[44]. Though the methods we describe here are specific to the larval sampling of *C. quinquefasciatus* breeding habitats in an urban situation, it may be applicable to other species of mosquitoes whose breeding habitats are similar in nature.

3.2.1. The sampling design:

The major breeding sources of *C. quinquefasciatus* are cess pits, cess pools, unused wells and drains in an urban situation[10,43]. The study area can be broadly classified into low, moderate and heavy breeding potential areas. Each of the above strata is comprised of substrata like villages. All the villages can be serially numbered and the habitats identified by house numbers or by giving some identification numbers to each type of habitat. For this purpose an initial survey may be carried out to enumerate the cess pits, cess pools, wells and drains. The form-1 may be completed for each habitat in each village. Since the breeding habitats may be widely scattered it is impossible to cover the entire area in each of these stratum. Alternatively a sample of habitats may be randomly selected in each stratum and dipper

or bucket samples are taken only from these habitats. For each type of habitat the form - 2 may be completed. An estimate of the mean of the population per fixed number of dips for each type of habitat can be calculated by the following formula:

$m = \Sigma N_i \mu_i / \Sigma N_i$ Where,

m_i = sample mean of ith stratum = $\Sigma x_{ij} / \Sigma n_i$

x_{ij} = Total number of larvae per fixed number of dips from jth unit (cess pits or wells or cess pools) of ith stratum.

n_i = Total number of units sampled in the ith stratum.

N_i = Total number of units in the ith stratum.

The variance of the mean 'm' is given by Cochran as,

variance of (m) = $\Sigma(W_i S_i / n_i)^2$

where, $W_i = N_i / \Sigma N_i$

$S_i^2 = [1/(n_i-1)][\Sigma(x_{ij} - m_i)^2]$

= Sample variance of ith stratum.

A comparison of different habitats can be made by taking the weighted average of estimated mean population in each habitat with weight being equal to the proportion of area actually breeding to total area of that habitat in each stratum[45].

FORM - 1.: Breeding habitats enumeration form.

Type of Habitat Date:

S.No.	Village code	Total No.. of habitats	Surface area (Sq.m)
..	...	..	..
5	10	120	500

FORM - 2.: Proforma for Larval Sampling

Village Code: Habitat: Date:

Habitat I.D.No.	No. of. dips. taken	No. of larvae in stage I	II	III	IV	P	Total Larvae
..	..	..	..	..	..	..	..
15	6	48	23	12	8	4	95

3.3. SAMPLING THE ADULT POPULATION:

As it is not possible to examine the entire mosquito population in a large area, a representative sample has to be collected over a fixed period of time. The purpose of sampling the adult mosquitoes can be broadly classified as follows:

1. To study the population dynamics of the vector.
2. To study the transmission dynamics of the parasite from vector to man.
3. To give feed back and evaluate the vector control programmes.

Adult mosquitoes can be sampled from the resting or biting population. Collection of resting or biting mosquitoes by hand catch provide more representative samples of the population. Resting or biting mosquitoes are collected by hand using either aspirators or a test tube for a fixed time period from the same houses or baits. The data obtained can be expressed as the average number of mosquitoes collected per man per hour as an index for comparing the seasonal fluctuation in the population. Though sampling the biting population would be ideal from epidemiological point of view, it is less practicable due to the following constraints:

1. Getting baits for mosquito collection.

2. If the population is of periodic in nature, then a biting sample should coincide with the peak biting hours. Since *C. quinquefasciatus* are known to be nocturnally periodic, a sample obtained from early or late hours may not provide a representative sample. Hence a whole night biting collection has to be made which is practically difficult.

Since the resting mosquitoes include a random mixture of fed as well as unfed mosquitoes and it is practically possible to collect mosquitoes during day time, it would not be biased if estimates are obtained from a sample of resting population. This could be true at least, if not all, for mosquito species with nocturnally biting rhythm. Further a whole night biting collection is useful only for determining or identifying the local vector species and their periodicity in nature.

3.3.1. Estimation of entomological Parameters:

The entomological parameters commonly used in the evaluation of control programmes are, the relative biting or resting densities, survival of the vector population, infection and infectivity rates, the annual infective biting rate (AIBR) and the annual or monthly transmission potential (ATP or MTP) of the vector population. WHO (1984) in its fourth report on lymphatic filariasis stated that "a comparison of these parameters, especially the AIBR and ATP, before and after control operation is valid, provided the sampling and dissection methods are standardized[1]". Some of these important parameters are described[2,3,5] below.

Annual Biting Rate (ABR):

This is the number of bites received by one man in one year. This parameter is estimated from direct 12 hours human bait landing catches carried out once in a fortnight or a month for a year.

Annual Infective Biting Rate (AIBR):

This is expressed as the number of infective mosquitoes biting a man in one year. The product of the infective rate in each catch and the estimated fortnightly or monthly biting rate gives the infective biting rate for that period. The AIBR is obtained by the summation of this parameter for a year (24 fortnights or 12 months accordingly).

Annual Transmission Potential (ATP):

It is the number of infective larvae that could have been received by one person in a year. This parameter is calculated using the following formula.

ATP = a X b X c where,

a = Estimated number of mosquitoes biting a man in one year (This is calculated as per man hour biting density X 12 hours X 365 days).

b = Proportion of mosquitoes infective (from biting collection only).

c = Number of infective larvae per infective mosquito.

Further, we describe some new methods for estimating the survival rate of adult mosquitoes and the relevance of currently available parameters for measuring the level of transmission.

3.3.1.1. The survival rate of vector:

The survival rate of vectors is an important component in determining the efficiency of transmission. This is the basic for determining the expected infective life span of the vector. One of the commonly used methods for estimating survival rate is based on the mean parous rate. Field collected mosquitoes are dissected and graded according to their parity status as parous or nulliparous etc. Then the survival rate (S) is calculated using the following formula:

S = nth root of proportion parous (P)

where

$$P = \text{Parous rate} = \frac{\text{Total parous}}{\text{No. dissected}}$$

n = the interval between two blood meals.

The standard error may be estimated by the following formula:

Standard error (S) = sqrt [S (1-S)/no.dissected]

The above formula makes the following assumptions:

1. The recruitment rate of nulliparous mosquitoes is constant between two time points of survey.

2. The population is stable., and

3. The duration between blood meals is constant for all stages of the mosquito.

However the recruitment rates and interval between blood meals are seldom constant in nature and consequently the changes in the proportion parous may bias the estimate of the survival rate. Based on the above reasonings new technique has been developed for populations which are rapidly changing over time[46]. These authors have expressed the dynamics of the biting mosquito population by two equations as below:

$$M_t = P_u T_{t-u}$$

$$T_t = N_t + M_t$$

where T_t is the total biting population on day t, composed of N_t nulliparous and M_t parous individuals. The proportion P_u is the survival rate per gonotrophic cycle, which has a mean duration of u days. Then an estimate of the P_u is obtained by the following formula,

$$P_u = \Sigma M_{ti} / \Sigma T_{ti-u}$$

Where M_{ti} is the parous sample in the total sample T_{ti} collected on day t_i, and there are m pairs of sampling occasions separated by u days. This method assumes that the time series M(t) and T(t-j), lagged by j days are correlated. Cross correlation coefficients are estimated with varying time lags and the lag at which the peak correlation occurs is the best estimate of the mean length of the cycle. The introduction of the time lag u is associated with the gonotrophic cycle of the biting population. This time lag solves the problem of varying recruitment rate, which biases the mean parous rate formula. However the integer u, the mean length of the cycle, is unknown and it, too must be estimated from the data.

3.3.1.2. Distribution of parasite in the vector mosquitoes.:

The distribution of microfilaria in the human population has been adequately described either by the negative binomial or lognormal probability functions[3,28,29]. However, the fitting of lognormal probability distribution to the observed microfilaria distribution did not have any biological explanation. Further modelling a discrete data (microfilaria counts) by a continuous probability distribution is not reasonable. Grenfell *et al.* (1990)[28] and Das *et al.*[29] have explained the biological mechanisms that generate the over dispersion of parasite in the human population and the consequence of describing this by a negative binomial distribution.

Das *et al. (1991)*[47] have recently observed that the distribution of parasite infection by vectors of lymphatic filariasis (as measured by mf counts) can be described by the same statistical model as that proposed for blood sampling mf by Grenfell *et al.(1990)*[28]. In particular, the mf count distribution in nulliparous mosquitoes is remarkably well described by the (zero-truncated) mixed Poisson distribution, which was used to fit the equivalent pattern of mf counts in humans.

The similarity of the observed vector and human distributions, therefore indicates either that counts from mosquito dissections reflect a higher sampling efficiency, or that there may be some concentration of parasites by the vector. The overdispersion analysis also confirms that parasite distributions become less overdispersed with parasite stage, reflecting a lower proportion of vectors with high parasite loads. This provides evidence of density dependence in the host-parasite relationship, either in parasite survival or parasite-induced host mortality.

4. CONCLUDING REMARKS

A series of important papers were published on reviewing quantitative techniques that are often used for analyzing data from filariasis control programmes[2,48-50]. The authors discussed and criticized the methods used for estimating the microfilaria density and periodicity until the end of 1960 and provided some other methods. However, none of the workers seem to have pointed out the bias while estimating parameters, which is the function of sampling techniques among others. Aikat *et al* emphasized the need for probability sampling and provided a stratified two-stage probability sampling design[51]. But their design did not take into account the factors that influence the estimation of parameters. However, from their study valid inferences were drawn to make use of it in the subsequent surveys. The sampling design proposed here takes care of practical constraints the authors have faced while evaluating a control programme in Pondicherry. The design was prepared by giving due consideration to the factors that would affect the estimation of parameters while evaluating any control programme.

The assessment of filariasis control programmes, either by chemotherapy or vector control, is based on clinical, parasitological, and entomological methods[1]. The parasitological assessment is mainly based on the microfilaria prevalence and intensity in a community. The degree of reduction is an indication of effectiveness of the control measures. This is particularly so in drug control programmes, which directly affect the fecundity of adult worms. However, the assessment based on this parameter in areas where only vector control measures were carried out, leads to paradoxical situations. This could be avoided by applying more sophisticated statistical techniques.

Though the understanding of parasite distribution in human host and its relationship with disease are well established, the knowledge on parasite distribution in the vector and its transmission dynamics to man and vice versa is at infant stage. The quantitative relationship between infection rates and intensities in the human population and the amount of transmission of the parasite to the development of microfilaria and of disease are essential for planning and monitoring of vector control programmes[52]. Undoubtedly these are the unresolved puzzles which require further immunological and statistical/mathematical techniques.

5. REFERENCES

1. World Health organization (1984). Fourth report of the WHO expert committee on filariasis: Lymphatic Filariasis. *Technical Report Series,702,* pp:1-112.

2. Mak, J.W. and Yong, H.S. (1986). *Control of Brugian filariasis.* Proceedings of the W.H.O. Regional seminar, *1-5th, July, 1985, Kuala Lampur,* pp:1-174.

3. Sasa, M. (1976). *Human Filariasis: A Global Survey of Epidemiology and Control.* University Park Press, Baltimore, London and Tokyo, pp:1-819.

4. Mak, J.W. (1987). In: *Filariasis.* Ciba Foundation Symposium, No.127. John Wiley & Sons Ltd., Chichester, UK, pp:5-14.

5. Anonymous (1988). *Filariasis:* A training module for middle level worker. *Misc. Publ. Vector Control Research Centre, 8,* pp:1-19.

6. Cochran, W.G. (1972). *Sampling Techniques.* Second edition, Wiley Eastern Private Ltd. New Delhi, pp:1-413.

7. Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey, Methods and Theory: Volume-I: Methods and Applications.* John Wiley & Sons Inc. New York, pp:1-638.

8. Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample, Survey, Methods and Theory: Volume-II:Theory.* John Wiley & Sons Inc. New York, pp:1-332.

9. Lwanga, S. (1978). Statistical principles of monitoring and surveillance in public health. *Bull. Wld. Hlth. Org., 56:* 713-722.

10 Rajagopalan, P.K. and Das, P.K. (1987). *The Pondicherry Project. On integrated disease vector control (Filariasis control demonstration project)* Vector Control Research Centre, Pondicherry, India, pp:1-164.

11. Krebs, C.J. (1989). *Ecological methodology.* Harper & Row Publishers, New York, pp:1-653.

12. Kahn, H.A. and Sepos, C.T. (1989). *Statistical Methods in Epidemiology.* Oxford University Press, New York, pp:1-292.

13. Armitage, P. and Berry, G. (1987). *Statistical Methods in Medical Research.* Blackwell Scientific Publications, Oxford, pp:1-559.

14. Kimura, E., Penaia, l. & Spears, G.F.S. (1985). Epidemiology of sub periodic bancroftian filariasis in Samoa: 8 years after control by mass treatment with diethylcarbamazine. *Bull. Wld. Hlth. Org., 63:4D 869-880.*

15. Pani, S.P., Krishnamoorthy, K., Rao, A.S. and Prathiba, J. (1990). Clinical manifestations in *Malayan filariasis* infection with special reference to Lymphoedema grading. *Indian J. Med. Res., 91:* 200-207.

16. Manson-Bahr, P.E.C. and Bell, D.R. (1987). In: *Mansons Tropical Diseases,* 19th Edn. The English Language Book Society and Bailliere Tindall, London. pp: 353-406.

17. Anonymous. (1985). Lymphatic Pathology and immuno-pathology in filariasis: Report of 12th meeting of the Scientific Working Group on filariasis. *WHO, TDR/FIL-SWG* (12)/85.3, pp: 1-38.

18. Pani, S.P., Balakrishnan, N., Srividya, A., Bundy, D.A.P. and Grenfell, B.T. (1990). Clinical Epidemiology of *Bancroftian filariasis:* effect of age & gender (submitted to *Trans. Roy. Soc. Trop. Med. Hyg.).*

19. Paniker, K.N. Krishnamoorthy, K., Sabesan, S., Prathiba, J. and Abidha. (1990).

Comparison of Effects of Mass Annual and Biannual single Dose therapy with Diethylcarbamazine (DEC) for the Control of *Malayan filariasis.* (Submitted to *J. Clin. Epidem).*

20. Avram Goldstein (1964). *BioStatistics:an introductory text.* The Macmillan company New York, pp:1-272.

21. Anderson, R.M. and May, R.M. (1978). Regulation and stability of host-parasite population interactions:I Regulatory processes. *Journal of Animal Ecology, 47:* 219-248.

22. Elliot, J.M. (1971). *Statistical Analysis of Samples of Benthic Invertebrates.* Scientific Publications, *25:* pp:1-143.

23. Anderson, R.M. and Gordon, D.M. (1982). Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology, 85:* 373-398.

24. Anne Keymer. (1982). Density-dependant mechanisms in the regulation of intestinal helminth populations. *Parasitology, 84:* 573-587.

25. Anderson, R.M. and May, R.M. (1982). Population dynamics of human helminth infections: Control by Chemotherapy. *Nature, 297:* 557-563.

26. Adjei, E.L., Barnes, A. & Lester, R.J.G. (1986). A method for estimating possible parasite related host mortality, illustrated using data from *Callitetrarhynchus gracilis (Cestoda:Trypanorhyncha)* in lizardfish *(Saurida spp.). Parasitology, 92:* 227-243.

27. Park, C.B. (1988). Microfilarial density distribution in the human population and its infectivity index for the mosquito population. *Parasitology, 96:* 265-271.

28. Grenfell, B.T., Das, P.K., Rajagopalan, P.K. and Bundy, D.A.P. (1990). Frequency distribution of microfilaria in humans: effects of infection prevalence and intensity. *Parasitology* (Accepted for publication, 1990).

29. Das, P.K., Manoharan, A., Srividya, A., Grenfell, B.T., Bundy, D.A.P. and Vanamail, P.(1990). Frequency distribution of *W.bancrofti* microfilariae in human populations and the relationships with age and sex. *Parasitology* (Accepted for publication, 1990).

30. Anderson, R.M. and May, R.M.(1985). Helminth infections of humans: mathematical models, population dynamics and control. *Advances in Parasitology, 24:* 1-101.

31. Muench, H. (1959). *Catalytic Models in Epidemiology.* Harvard University Press:Cambridge.

32. Hairston, N.A. and Jackowski, L.A. (1968). Analysis of the *W.bancrofti* population in people of Western Samoa. *Bull. Wld. Hlth. Org., 38:* 29-59.

33. Vanamail, P., Subramanian, S., Das, P.K., Pani, S.P., Rajagopalan, P.K., Bundy, D.A.P. and Grenfell, B.T. (1989). Estimation of the Age-specific rates of Acquisition and loss of *Wuchereria bancrofti* infection. *Trans. Roy. Soc. Trop. Med. Hyg., 83:* 689-693.

34. Rajagopalan, P.K., Das, P.K., Subramanian, S., Vanamail, P. and Ramaiah, K D. (1989). Bancroftian filariasis in Pondicherry,South India: I.Pre-control epidemiological observations. *Epidemiology & Infection, 103:* 685-692.

35. Srividya, A., Pani, S.P., Rajagopalan, P.K., Bundy, D.A.P. and Grenfell, B.T. (1990). The Dynamics of Infection and Disease in *Bancroftian filariasis.* (submitted to Trans. Roy. Soc. Trop. Med. Hyg.).

36. Ottesen, E.A. (1984). Immunological aspects of lymphatic filariasis and onchocerciasis in man. *Trans. Roy. Soc. Trop. Med. Hyg., 78* (suppl): 9-18.

37. Ottesen, E.A., Mendell, N.R., Macqueen, J.M., Weller, P.F., Amos, D.B. & Ward, F.E. (1981). Familial predisposition to filarial infection - not linked to HLA-A or -B locus specificities. *Acta Tropica, 38:* 205-216.

38. Rao, C.K., Krishna Rao, Ch. & Narasimham, M.V.V.L. (1981). Bancroftian filariasis prevalence by family size in East Godavari district (Andhra Pradesh). *J. Commun. Dis., 13:* 281-283.

39. Vanamail, P., Subramanian, S., Das, P.K., Pani, S.P. & Bundy, D.A.P. (1989). Familial clustering in *Wuchereria bancrofti* infection. *Tropical Biomedicine, 6:* 67-71.

40. Southwood, T.R.E. (1978). *Ecological Methods.* Second edition. Chapman & Hall, London, pp:1-524.

41. Service, M. (1976). *Mosquito Ecology: field sampling technique:* Applied Science Publishers, London, pp:1-583.

42. Rajagopalan, P.K., Das, P.K. and Velayudhan, R. *Culex quinquefasciatus* SAY, 1823: with special reference to Asia. *Training and Information Guide, WHO. Vector Control Series,* pp:1-150.(Submitted).

43. Menon, P.K.B. and Rajagopalan, P.K. (1980). Relative importance of different types of breeding habitats in contributing to the population of *Culex pipiens fatigans* in Pondicherry. *Indian J. Med. Res., 71:* 725-733.

44. Poti, S.J. (1974). *A note on Sampling Methods.* A report of the WHO consultant to *WHO/ICMR Research Unit on Genetic Control of Mosquitoes,* New Delhi, 2nd August. pp:1-19.(Mimiographed document).

45. Cairncross, S., Rajavel, A.R., Vanamail, P., Subramanian, S., Paily, K.P., Ramaiah, K.D., Amalraj, D., Mariappan,T. and Srinivasan, R. (1988). Engineering, mosquitoes and filariasis: a case report. *J. Trop. Med. Hyg., 91:* 101-106.

46. Birely, M.H. and Rajagopalan, P.K. (1981). Estimation of survival and biting rates of *C.quinquefasciatus* (Diptera:Culicidae). *J. Med. Ento., 18:* 181-186.

47. Das, P.K., Subramanian, S., Manoharan, A., Ramaiah, K.D., Vanamail, P., Grenfell, B.T. and Bundy, D.A.P. Frequency distribution of *Wuchereria bancrofti* infection in vector host.(Sent for publication).

48. Anonymous. Operational Manual. *National Filaria Control Programme* (1984). NMEP, Sham Nath Marg, New Delhi, India, pp:1-152.

49. Sasa, M. (1967). Microfilaria survey methods and analysis of survey data in filariasis control programmes. *Bull. Wld. Hlth. Org., 37:* 629-650.

50. Sasa, M. and Tanaka, H. (1972). Studies on the methods for comparison and classification of the microfilarial periodicity. *South East Asian Journal of Tropical Medicine and Public Health, 3:* 518.

51. Aikat, T.K., Tapansen and Singh, N.N. (1974). A sample design for filariasis survey and the results of its field trial. *J. Commu. Dis., 6:* 284-289.

52. Southgate, B.A. (1984). Recent Advances in the epidemiology and control of filarial infections including entomological aspects of transmission. *Trans. Roy. Soc. Trop. Med. Hyg., 78* (suppl): 19-28.